

METHODOLOGY ARTICLE

Open Access

# BayMiR: inferring evidence for endogenous miRNA-induced gene repression from mRNA expression profiles

Hossein Radfar<sup>1</sup>, Willy Wong<sup>1</sup> and Quaid Morris<sup>1,2,3,4\*</sup>

## Abstract

**Background:** Popular miRNA target prediction techniques use sequence features to determine the functional miRNA target sites. These techniques commonly ignore the cellular conditions in which miRNAs interact with their targets *in vivo*. Gene expression data are rich resources that can complement sequence features to take into account the context dependency of miRNAs.

**Results:** We introduce BayMiR, a new computational method, that predicts the functionality of potential miRNA target sites using the activity level of the miRNAs inferred from genome-wide mRNA expression profiles. We also found that mRNA expression variation can be used as another predictor of functional miRNA targets. We benchmarked BayMiR, the expression variation, Cometa, and the TargetScan “context scores” on two tasks: predicting independently validated miRNA targets and predicting the decrease in mRNA abundance in miRNA overexpression assays. BayMiR performed better than all other methods in both benchmarks and, surprisingly, the variation index performed better than Cometa and some individual determinants of the TargetScan context scores. Furthermore, BayMiR predicted miRNA target sets are more consistently annotated with GO and KEGG terms than similar sized random subsets of genes with conserved miRNA seed regions. BayMiR gives higher scores to target sites residing near the poly(A) tail which strongly favors mRNA degradation using poly(A) shortening. Our work also suggests that modeling multiplicative interactions among miRNAs is important to predict endogenous mRNA targets.

**Conclusions:** We develop a new computational method for predicting the target mRNAs of miRNAs. BayMiR applies a large number of mRNA expression profiles and successfully identifies the mRNA targets and miRNA activities without using miRNA expression data. The BayMiR package is publicly available and can be readily applied to any mRNA expression data sets.

## Background

MicroRNAs are short (21-25 nt) non-coding RNAs that repress the expression of their direct targets [1-4]. Primary miRNAs (pri-miRNAs) are transcribed from intra/intergenic genomic loci and cleaved by Drosha to form approximately 70-nt hairpin precursors (called pre-miRNAs) that are subsequently cleaved by the RNase III enzyme, Dicer, to generate miRNA duplexes [5]. One strand of the duplex, the mature miRNA, is loaded into

the RNA-induced silencing complex (RISC) [6] and guides it to recognize mRNA targets through partial base pairing with the 3' UTRs of targets [7].

The presence of target sites with perfect complementarity to the seed region of miRNAs is a strong predictor of targeting but perfect complementarity is neither sufficient nor necessary [7-10]. Many other determinants have been proposed to specify efficient mRNA-miRNA duplexes including: AU composition flanking target sites [8], thermodynamic stability of binding sites [11], evolutionary conservation of the seed [12-14], secondary structure accessibility [6,15-17], target-site abundance [18,19], seed-pairing stability [18], 3' pairing contribution [8], loop in position 9-12 of miRNA-mRNA hybrids [10], and the binding location in the 3' UTR [8,17]. Due to the

\*Correspondence: quaid.morris@utoronto.ca

<sup>1</sup>Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada

<sup>2</sup>Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

Full list of author information is available at the end of the article

limited number of validated miRNA targets, the exact specificity and sensitivity of current determinants are unclear [20-23]; however, estimates of precision of these determinants, alone or together, are typically reported to be about 50% at a sensitivity of 6-12% [24,25], suggesting that sequence-based prediction methods are not fully capturing miRNA target preferences.

In mammals, it is estimated that miRNAs primarily and dominantly repress the steady-state expression level of their targets [26-34]. Therefore, down-regulation of an mRNA's expression when the miRNA is active is evidence of a functional target site on the gene *in vivo*. Although numerous methods have been introduced to incorporate mRNA and miRNA expression data into miRNA target predictions, existing methods either require paired miRNA-mRNA data [35-48], have only been tested in miRNA transfection assays [28,29,49], or do not consider the combinatorial impact of multiple miRNAs on mRNA expression [50,51].

In this paper, we introduce two new mRNA-miRNA scoring schemes by incorporating genome-wide measures of mRNA expression in target prediction. Neither of these scoring schemes requires miRNA expression data, so can be applied to vast amount of publicly available mRNA expression databases. The first scoring scheme identifies the impact of a miRNA in repressing an mRNA in presence of other targeting miRNAs, cellular activities, and under a wide range of endogenous conditions. This scheme (hereafter called the BayMiR score) is obtained using BayMiR, a sparse Bayesian linear regression model, in which the decrease in expression levels of an mRNA across different conditions is explained in terms of the activity of miRNAs that have conserved target site matches in the 3' UTR of the transcript. BayMiR infers miRNA activity levels based on the expression profiles of its putative targets (predicted on the basis of conserved seed matches) and then it refines these target predictions using the regression model. We also found that expression variability is significantly higher among mRNAs with more miRNA target sites and, furthermore, that it can be used to identify more likely targets. Accordingly, we used the variance of gene expression levels across a wide range of samples including different cell types, cell lines, and disease/healthy tissues as another mRNA-miRNA scoring scheme. These scores are called "gene variation" index.

BayMiR analysis was conducted on 1,539 human miRNAs and the expression levels of 13,303 genes measured on 5,372 microarray experiments and predicts that approximately 60% of miRNA-mRNA duplexes with matched conserved targets sites have detectable down-regulation signal on gene expression. We evaluated and compared the efficacy of the proposed scores with eight TargetScan scores (a collection of most important sequence based features) as well as Cometa scores

(an mRNA expression based miRNA target prediction method) using over-expression miRNAs experiments, validated targets, and GO and KEGG enrichment analysis. Using these benchmarks, we found the BayMiR scores consistently outperform both the sequence and expression scores and identify to what extent down-regulated genes on a global set of microarrays are under control of miRNAs.

## Results

### BayMiR method

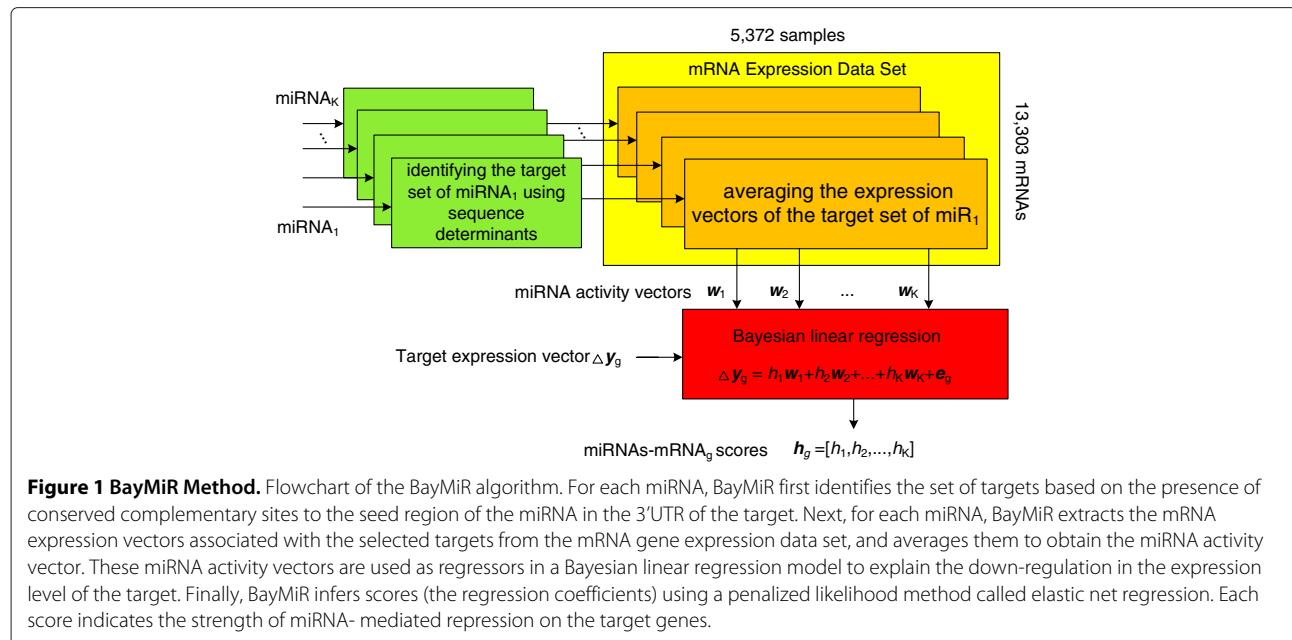
BayMiR (Figure 1) calculates the degree to which mRNA down-regulation inferred from a large set of microarrays can be explained by inferred miRNA activity. BayMiR makes this prediction by integrating sequence and expression evidence. Because many targets are under the control of multiple miRNAs [20,46,52,53], BayMiR applies a linear model that relates the target expression vector (measured variable) to a weighted combination of the miRNA activity vectors (regressor variables). BayMiR infers the activity vector of a given miRNA by averaging the normalized expression vectors of its predicted mRNA targets based on sequence-based prediction methods. These miRNA activity vectors are then used as regressors in a Bayesian linear regression model of the "down-regulation" expression vector of each mRNA. The resulting regression coefficients of each miRNA are interpreted as the strength of miRNA-mediated repression of the target mRNA.

We also considered the variability in gene expression of a target mRNA as a determinant to distinguish functional and non-functional targets of a given miRNA. The gene variation index for each mRNA is computed as the variance of gene expression levels across all samples.

Each expression vector consists of the transcriptional abundance of the target in one of 392 biological samples collected from 5,372 microarray experiments. We determine the coefficients of the regression model using a penalized likelihood approach called elastic net regression [54] (see Methods) modified to assign only positive coefficients. By using this regression model, each sequence-predicted miRNA-mRNA interaction is assigned one coefficient; this coefficient represents how much the inferred activity profile of that miRNA contributes to predicting that mRNA's "down-regulation" profile (see Methods) when considering the activity profiles of all other miRNAs predicted to target the mRNA. We call these coefficients "BayMiR scores" and interpret a zero BayMiR score as representing a lack of evidence in the expression data for regulation of the mRNA by that miRNA.

### BayMiR identifies highly repressed targets on miRNA over-expression assays

To evaluate whether the BayMiR scores reflect the strength of miRNA-mediated repression of mRNA tar-



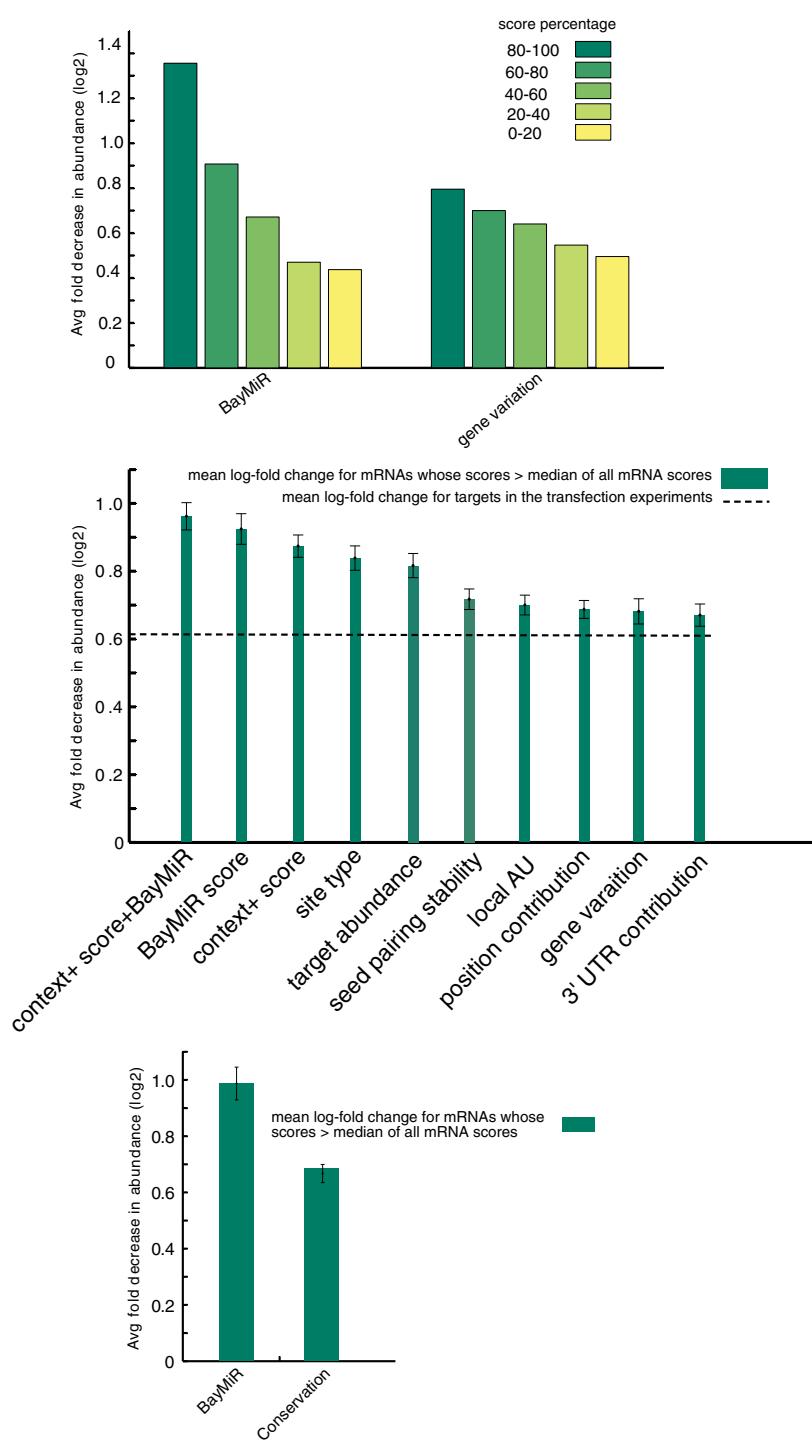
**Figure 1 BayMiR Method.** Flowchart of the BayMiR algorithm. For each miRNA, BayMiR first identifies the set of targets based on the presence of conserved complementary sites to the seed region of the miRNA in the 3'UTR of the target. Next, for each miRNA, BayMiR extracts the mRNA expression vectors associated with the selected targets from the mRNA gene expression data set, and averages them to obtain the miRNA activity vector. These miRNA activity vectors are used as regressors in a Bayesian linear regression model to explain the down-regulation in the expression level of the target. Finally, BayMiR infers scores (the regression coefficients) using a penalized likelihood method called elastic net regression. Each score indicates the strength of miRNA-mediated repression on the target genes.

gets, we measured the consistency between the BayMiR scores and relative down-regulation of targets in a set of miRNA over-expression experiments. One expects high scoring targets to be down-regulated more in miRNA over-expression experiments. We note that a similar metric has previously been used to evaluate the efficiency of TargetScan scores [8,18], and that this set of miRNA over-expression assays were not used in BayMiR to obtain the scores; thus, we are not influencing the results of our evaluation by either selecting bias metrics or by evaluating our model on the training data. We downloaded the data collected by Khan et. al [34] in which 23 miRNAs were transfected into seven different cell types and the log-fold change of the expression levels of mRNAs were measured. To examine that the degree to which our scores can predict the log-fold change of mRNAs in the miRNA over-expression arrays, for each score, we binned mRNAs into five bins based on their scores and computed the mean of mRNA log-fold changes in each bin. We observed that negative log-fold repression levels decrease consistently as scores decrease for both determinants (Figure 2.(top)). In total, 3,867 out of 10,125 mRNAs are down-regulated in the miRNAs over-expression experiments. We then asked if our scoring schemes can detect repressed targets better than the individual components of the TargetScan context score [8]. When comparing negative mean log-fold changes for messages whose scores were greater than the median score for the corresponding miRNA, BayMiR scores outperforms all TargetScan scores, even the context+score which is a combination of all individual TargetScan scores (Figure 2.(middle)). In addition, when we combined BayMiR scores and the Tar-

getScan context+score the performance further improved (Wilcoxon-Mann-Whitney test:  $P < 0.001$ ), indicating that BayMiR can augment the TargetScan scoring system to further improve the performance. Target site conservation is another scoring scheme used by TargetScan, so we also compared BayMiR scores with conservation scores for all conserved target sites of all conserved miRNA families and found similar improvements (Figure 2.(bottom)). Our analysis also shows that the gene variation score was a better predictor of log-fold change than seed pairing stability, relative location of seed match in the 3' UTR, and target abundance; however, it is worse than the other components of the context score on this assay (Figure 2(middle)).

#### High-scoring BayMiR targets are enriched for validated targets

To test whether the set of experimentally validated targets are enriched among high-scoring BayMiR targets, we measured the significance of overlap between the targets with scores greater than the median and the experimentally validated targets retrieved from TarBase [55]. Enrichment using the hyper-geometric test showed that the validated targets are enriched in the sets of high-scoring genes both for BayMiR and gene variation predicted targets,  $P < 10^{-5}$  and  $P < 10^{-4}$  respectively. A cumulative distribution analysis is also shown in Additional file 1: Figure S1. Number of TarBase validated human targets at mRNA level is 491; number of validated targets with conserved target site is 279 and BayMiR predicts 203 of these conserved validated targets (72.8%). Together these observations support that the hypothesis that repressed targets



**Figure 2 BayMiR performance in the miRNA over-expression experiments.** (top) mRNAs in the over-expression miRNA assays are grouped into five bins based on their BayMiR and gene variation scores; the mean log-fold change of the mRNAs in each bin is plotted in as a bar. There are two groups of bars; the left- and right-hand groups correspond to BayMiR and gene variation, respectively. (middle) Comparing BayMiR and gene variation scores with seven sequence scores from TargetScan. Each bar represents the negative mean log-fold change for mRNAs whose scores are greater than the median of all mRNA scores for the selected determinant in the miRNA over-expression assays. The most left-hand group is obtained by combining the context+ scores with BayMiR scores. The dashed line shows the mean log-fold change for all targets in the miRNA over-expression assays (bottom) Comparing BayMiR scores with the conservation scores as measured by TargetScan. The conservation scores are given only for the targets with conserved target sites complementary to the seed regions of the conserved miRNA families. Error bars indicate 95% confidence intervals for the estimated means.

under the endogenous conditions are more likely to be functional targets.

#### **BayMiR predicts miRNA-induced repression better than Cometa**

Next, we used the same evaluation strategy to compare BayMiR scores with an mRNA-miRNA scoring method which also uses large-scale gene expression data. Recently, Gennarino et al. [50] showed that the target set of a miRNA tend to be co-expressed and based on this property they proposed Cometa, a computational method that scores each sequence-based miRNA target prediction based on how correlated it is with other predicted targets of the miRNA. Examining the down-regulated targets on the miRNA over-expression assays shows that negative mean log-fold expression changes for targets selected by our scoring schemes are significantly higher than those selected by Cometa scores ( $P < 10^{-40}$ , Additional file 2: Figure S2). Moreover, our methods' high scoring targets are significantly more down-regulated compared to Cometa high scoring targets ( $P < 10^{-60}$  Figure 3) on the over-expression assays. Although Cometa targets are also enriched for validated targets, this enrichment is smaller than BayMiR scoring targets ( $P < 0.01$  v.s.  $P < 10^{-5}$ ).

#### **BayMiR target sets have more consistent GO-BP and KEGG annotations**

Many miRNAs participate in the coordinate regulation of biological processes [56]; as such, we should expect that, in general, better target prediction methods would generate miRNA target sets that have higher enrichment [57]. To test whether BayMiR predicted targets are more consistently annotated with GO and KEGG terms than TargetScan targets, we used Fisher's exact test with an FDR multiple test correction (see Methods) to score the enrichment of 1,233 GO-BP terms and 259 KEGG pathways within the target sets of each of 1,264 miRNA families. We found a nearly three-fold increase in enriched terms and pathways ( $FDR < 0.1$ ) within BayMiR-predicted target sets compared to equally-sized random subsets of TargetScan (31,976 vs 11,890,  $P < 10^{-200}$ ).

Examination of the enriched GO-BP terms and KEGG pathways revealed a wide diversity of biological processes regulated by miRNAs (Additional file 3: Table S1,  $FDR < 0.1$  and Additional file 4: Table S2,  $FDR < 0.1$ ). We found that 35 % of miRNAs that have BayMiR target sets are enriched for the GO term "regulation of expression" suggesting that miRNAs have substantial influence in gene regulation through their control of other gene regulators.

We also searched for miRNAs with known functions among the miRNAs enriched in our pathway analysis. A list of miRNAs with experimentally supported functions

among their enriched pathways are given in Additional file 5: Table S3. Notably the miR-17 family is frequently seen in the list. This family has been extensively studied and shown to play an important role in many cancer-related processes and pathways [58,59], and references in Additional file 5: Table S3.

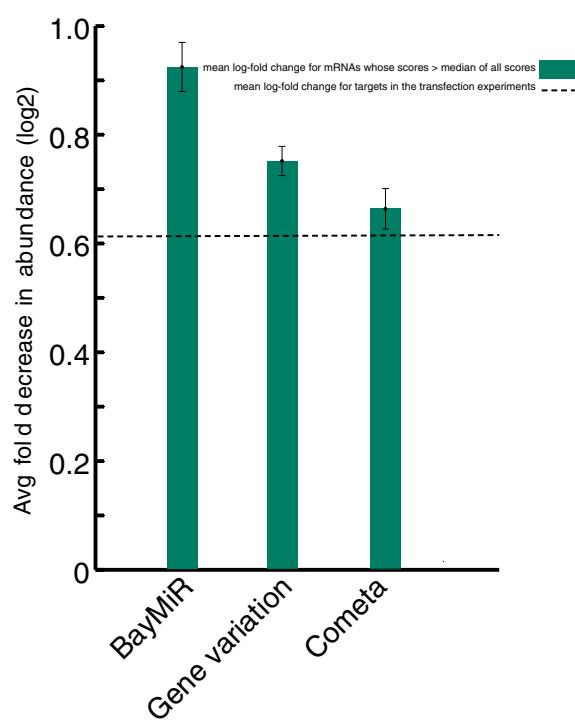
When we examined the mRNAs in KEGG pathways targeted by miRNAs, we found that although there are extensive co-regulation of mRNAs by multiple miRNAs, a handful of miRNAs appeared to be responsible for most of the regulation. For example, in the WNT signaling pathway, five miRNAs target 32 out of 46 genes predicted to be targeted by any of the 45 miRNAs with targets in this pathway (Figure 4). Similarly, the 106 genes in "Pathways in cancer" are targeted by 83 miRNAs but only 10 of these miRNAs collectively target more than 75% these genes (Additional file 6: Figure S3). Although some of this consolidation of targeting can be explained with a large variability in number of mRNA targets per miRNA, there is significantly more consolidation than we would expect by chance (Figure 5,  $P < 10^{-19}$ ) These observations suggest that important miRNA regulators of specific biological processes can be identified *in silico* through gene set enrichment analysis of BayMiR target sets.

#### **miRNA activity and expression profiles are significantly correlated**

To test if miRNA activities obtained using the BayMiR procedure are correlated with the miRNA expression profiles, we downloaded the miRNA expression data from the mimiRNA repository [60] and computed the correlation between matched activity and expression vectors. After excluding miRNA expression data that are not consistent across multiple resources (according to  $P > 0.05$  reported in the mimiRNA resource) and mapping the biological samples of the miRNA expression data to our biological groups we obtained paired matches for 48 miRNAs. Interestingly, we found that 96% of the pairs (46 out 48) have the Pearson correlation coefficients greater than 0.35 compared to 4% positive correlation obtained from a similar analysis but with the permuted activity vectors ( $P < 0.05$  and Additional file 7: Table S4). This correlation analysis shows that miRNA activities inferred from the mean of inverse expression of their targets are highly correlated with expression data for those miRNAs.

#### **mRNAs harboring miRNA target sites near the both ends of the 3' UTR have higher endogenous down-regulation signals**

To investigate any association between endogenous target repression scores provided by BayMiR and sequence and



**Figure 3 Comparing BayMiR and Cometa.** BayMiR high scoring targets are more down-regulated in miRNA over-expression assays than Cometa high scoring targets. Each bar represents the mean of negative log-fold change after miRNA over-expression for genes with scores greater than median.

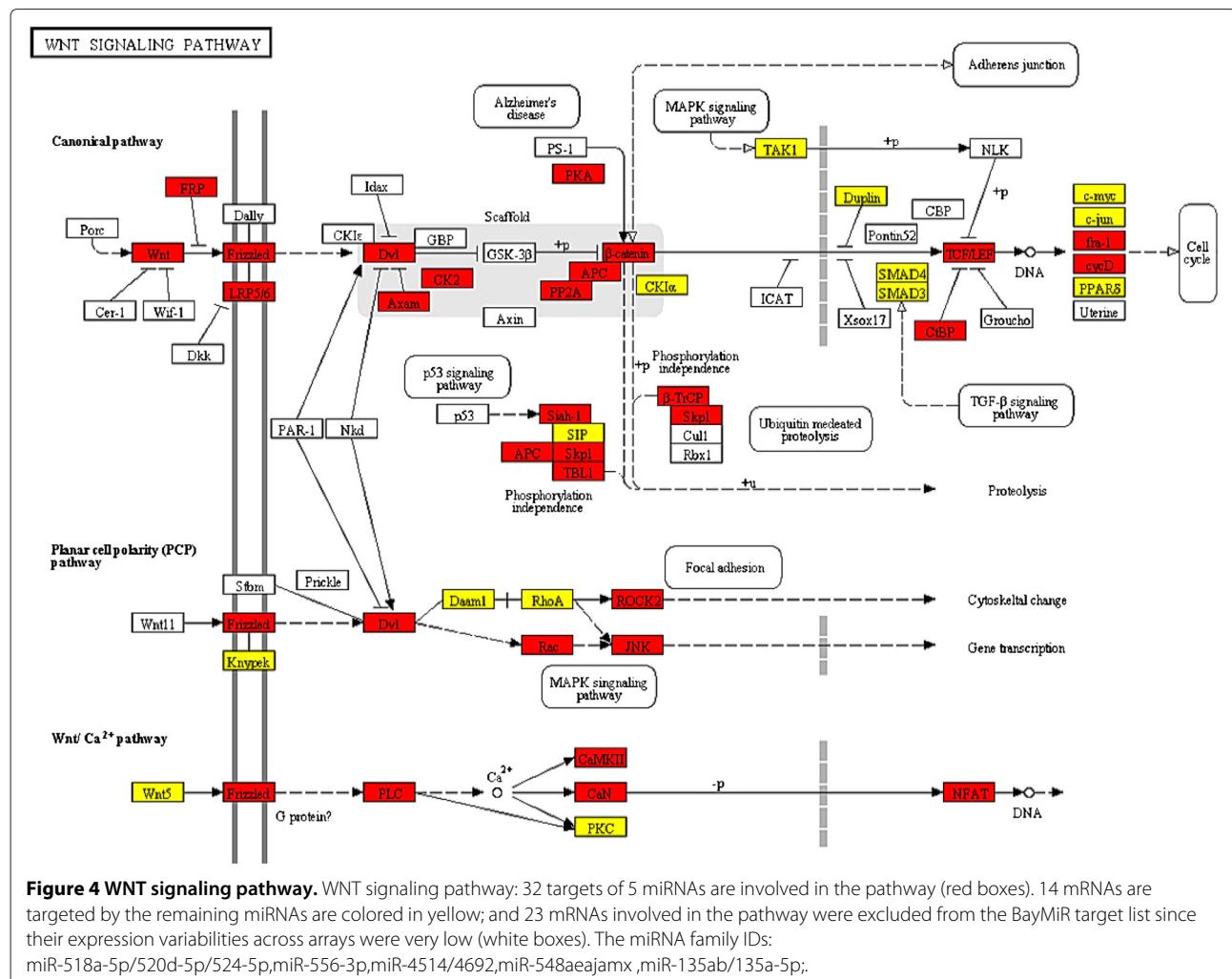
gene variation determinants, we measured the correlation between the scores of all paired determinants (Figure 6). The heat map shows that BayMiR scores correlate most highly with the position contribution scores. In addition, when we ranked all mRNA-miRNA pairs based on their BayMiR scores, the top 50 percentile of the ranked list have higher position contribution scores than the bottom 50 percentile ( $P < 10^{-200}$ , Wilcoxon-Mann-Whitney test and Additional file 8: Figure S7). The position contribution scores provide estimate of expected repression in terms of the distance of target sites from the both end of the 3' UTR; target sites near to the ORF or the poly(A) tail are more effective [8] and more conserved than those in the middle of the 3' UTR [12]. To further investigate this, we located 1,567,294 conserved target sites matched to the seed region of 1,032 miRNAs on the 3' UTR of 17,840 mRNAs. The start position of each target site was divided by the length of the 3' UTR to obtain the relative position of miRNAs on the 3' UTRs, denoted by  $0 < L_{miRNA} < 1$ . We found that target sites located on the both end of 3' UTRs ( $L_{miRNA} < 0.25$  or  $L_{miRNA} > 0.75$ ) are assigned higher BayMiR scores than those on the middle ( $P < 10^{-200}$ , Wilcoxon-Mann-Whitney test). Furthermore, we found that target sites located in the terminus close to the poly(A) tail ( $L_{miRNA} > 0.75$ ) are assigned higher BayMiR scores than to those located on

the other terminus ( $L_{miRNA} < 0.25$ ,  $P < 10^{-5}$ , Wilcoxon-Mann-Whitney test). Poly(A) shortening is known as one of the mechanisms of mRNA degradation; this mechanism strongly favors the preference of miRNA target sites near the end of 3'UTR close to the poly(A) tail to recruits mRNA deadenylase complexes [61]. Together these lines of evidence underline the importance of target site position in miRNA targeting.

BayMiR scores are also highly correlated with gene variation scores suggesting that mRNAs with high expression variability are under selective pressure to be miRNA targets.

## Discussion

Large-scale mRNA expression profiling datasets provide a rich resource to study the regulatory impact of miRNAs. Here, we showed that the impact of miRNAs on targets is detectable in normal tissue and unperturbed cell line data. Given a list of miRNAs with partial complementarity to a particular mRNA, our computational technique, BayMiR, scores the relative regulatory impact of the miRNA among other predicting targeting miRNAs. We showed that BayMiR estimates of miRNA regulatory impact better reflect independent measures of this impact than the TargetScan context scores; furthermore, we showed that the context scores and BayMiR

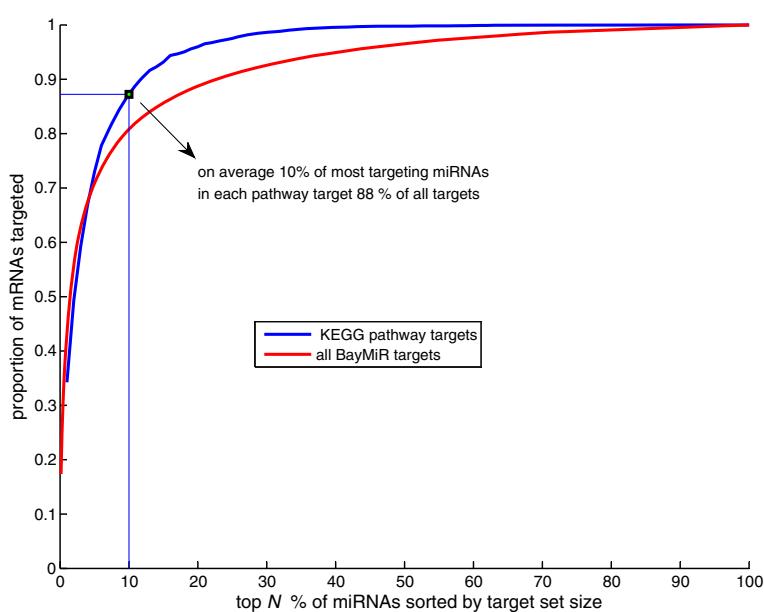


can be combined to generate even better estimates. We also demonstrated that the miRNA activity vectors that we infer from mRNA expression data are well-correlated with the measured expression levels of these miRNAs.

BayMiR has several features that make it particularly useful for estimating the potential regulatory impact of a miRNA. BayMiR models the combinatorial effect of multiple regulatory miRNAs on a single target which is critical, as most mRNAs are likely to be targeted by multiple miRNAs (Additional file 9: Figure S4). BayMiR is fast; its runtime is less than a minute in the current version, so is easily applied to a subset of or all available gene expression data. Because BayMiR estimates the activity of miRNAs based on mRNA expression data, there is no need for matching miRNA expression profiles. As such, BayMiR predictions can be easily extended when new miRNAs are found and the current version of BayMiR incorporates all miRNAs retrieved from the latest release of miRBase (v.19).

Combinatorial regulation by multiple miRNAs has been described for particular mRNAs [8,62] and is likely to play a large role in mRNA expression regulation [46]. Indeed, human 3' UTRs contain conserved seed matches for on average 33 of miRNAs (median = 16) (Additional file 9: Figure S4). This combinatorial regulation may explain the observations that inverse correlation under endogenous condition between miRNA and mRNA expression does not provide strong and consistent evidence of targeting [60,63] and that the impact of miRNA regulation on mRNA levels can only be seen within the context of other miRNA regulations [46,63]. Additional file 10: Figure S5 shows a toy example where combinatorial regulation masks inverse correlation between miRNA regulators and their targets.

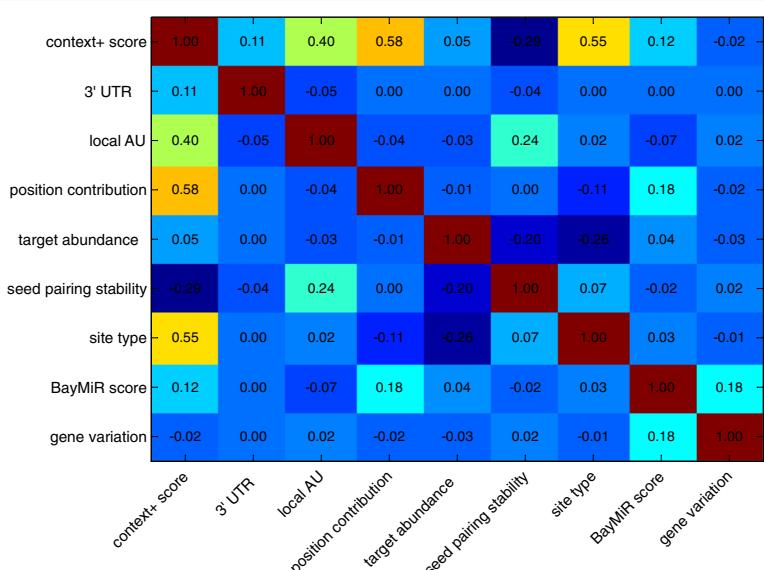
There are a large number of other methods [49-51,63-72] that infer either miRNA activity or predict miRNA targets based on the expression levels of their sequence-predicted targets, however, no method both infers miRNA activity and predicts miRNA targets while considering the impact



**Figure 5 miRNA targeting.** A small percent of all targeting miRNAs collectively target a large portion of miRNA targets. The figure shows two cumulative distributions. (red) Proportion of set of all mRNAs with BayMiR targets that are covered by union of the target sets of the top  $N\%$  miRNAs (sorted by number of targets) where  $N$  increases along the x-axis. (blue) Average of cumulative distributions for all enriched KEGG pathways ( $n = 108$ ), where each distribution was created as per the red line but restricted to targeted mRNA associated with the KEGG pathways, and their targeting miRNAs.

of other miRNAs. For example, Cometa attempts to predict miRNA targets, by identifying tight, co-expressed clusters of sequence-predicted targets [50]; however it doesn't account for combinatorial regulation by multiple miRNAs and provides no estimate of miRNA activity.

Other methods such as Sylamer [49], and a number of web-based applications [66–68], identify miRNA seed regions that significantly enriched in the 3' UTRs of down-regulated transcripts as a way of assessing miRNA activity level in a tissue. However, the performance



**Figure 6 Correlation between different determinants.** The heat map shows the Pearson correlation coefficients between each pair of nine determinants. A "0.00" correlation is shown for correlation coefficients that are not significantly different from 0 ( $P > 0.05$ ).

of Sylamer when applied to endogenous gene expression data is unclear. In addition, it does not take into account multiple targeting effect of miRNAs and has not been used to score the individual miRNA-mRNA pairs. Other methods use paired miRNA-mRNA expression patterns to augment sequence-based target prediction [35-48]. These methods typically require paired miRNA and mRNA measurements in a large number of samples to generate reliable predictions. This type of paired expression data is however rare and unavailable for some miRNAs [73]. On the other hand, there is very large amount of mRNA expression data available for BayMiR. Two intronic miRNA target prediction methods, InMiR and Hoctar [51,63] predict the intronic miRNA targets using the expression levels of their host genes, and subsequently can also incorporate large mRNA expression data. However, these methods can only be applied to intronic miRNAs and only to those miRNAs whose host gene expression is a good surrogate for their activity. Many host gene expression levels are not good surrogates [63,74-76].

Our analysis also reveals that mRNAs with more target sites have higher expression variation when compared to a random subset of genes, and expression variance consistently increases as number of target sites do ( $P < 10^{-33}$ , Additional file 11: Figure S6). These observations suggest that mRNAs with highly variable expression levels are much more likely to be regulated by miRNAs; our finding is consistent with recent reports that genes regulated by miRNAs have higher expression variability at among humans and between human and other primate species [77].

miRNA transfection experiments have suggested that the degree of mRNA repression induced by two seeds is equivalent to the product of repression induced by the seeds individually [8]. We have observed a similar effect. The version of BayMiR described here implicitly assumes multiplicative interactions because it log-transforms the mRNA expression levels before performing regression. Applying BayMiR to non-transformed expression levels assumes additive interactions and this version of BayMiR performs much worse in our benchmarks (data not shown).

In this paper, we introduced BayMiR and demonstrated its merits when compared to two the state-of-the-art miRNA computational prediction methods. BayMiR applies a more relevant biological model and uses a large collection gene expression data to decipher the impact of miRNAs on gene expression data. We measured this impact in terms of endogenous target repression scores for about half a million miRNA-mRNA duplexes. This new scoring strategy can be used alone or along with other sequence determinants to predict functional miRNA-mRNA interactions.

## Methods

### BayMiR model

BayMiR applies the following linear model to relate the changes in the log-transformed expression level of mRNAs to the activity level of miRNAs:

$$\Delta \mathbf{y}^i = \mathbf{W}_{M \times K} \mathbf{h}^i + \boldsymbol{\epsilon}_{M \times 1}$$

where  $\Delta \mathbf{y}^i \in \mathbb{R}^M$  denote the change in the expression level of the  $i$ th mRNA measured across  $M$  samples and is obtained by subtracting the mean from  $\mathbf{y}^i$ ;  $\mathbf{W} = [\mathbf{w}_{m,k}]_{M \times K}$  denote the activity levels of  $K$  miRNAs across  $M$  samples, and each element of  $\mathbf{h}^i \in \mathbb{R}^{+K}$  represents the contribution of the corresponding miRNA in down-regulating the expression of the  $i$ th mRNA;  $\boldsymbol{\epsilon}$  models error. In our problem  $K = 1,252$ ;  $M = 369$  and  $i = 1, \dots, 13,000$ .

In this linear equation,  $\Delta \mathbf{y}^i$  and  $\mathbf{W}$  and are observed;  $\mathbf{h}^i$  is the desired unknown variable. BayMiR infers  $\mathbf{h}$  by maximizing its posterior probability of  $\mathbf{h}$  given  $\Delta \mathbf{y}$  and  $\mathbf{W}$ :

$$\hat{\mathbf{h}} = \arg \max p(\mathbf{h} | \Delta \mathbf{y}, \mathbf{W}).$$

This inference problem can be written in form of a penalized linear regression optimization given by:

$$\begin{aligned} \hat{\mathbf{h}} = \arg \min & \sum_m (\Delta y_m - \mathbf{w}_{m,:} \mathbf{h})^2 + \lambda_1 \sum_k h_k \\ & + \lambda_2 \sum_k h_k^2 \text{ subject to: } h_k \geq 0 \forall k \end{aligned} \quad (1)$$

where  $\lambda_i$ s are two tuning parameters and  $\mathbf{w}_{m,:}$  is a row vector representing the expression activity of miRNAs in the  $m$ th sample. We solved this optimization using the coordinate-descent method [54] in which, the objective function is partially optimized with respect to each individual coefficient in an iterative manner given by

$$h_j = \frac{S\left(\sum_{m=1}^M (\Delta y_m - \sum_{k \neq j} w_{m,k} h_k) w_{mj}^n, \lambda_1\right)}{\sum_{m=1}^M w_{mj}^{n2} + \lambda_2} \quad (2)$$

where  $S(x, t)$  is the soft threshold operator defined as  $sign(x)(|x| - t)_+$  where  $(y)_+ = 0$  if  $y < 0$  and  $(y)_+ = y$  if  $y \geq 0$  [78].

Since miRNA and target mRNA expression data are anti-correlated [79], for each miRNA, BayMiR uses the negative mean of target expression levels as an estimate of the activity level of the miRNA as follows:

$$\begin{aligned} \mathbf{w}_k = -\frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{y}_i \text{ where} \\ N_k : \text{number of target genes for } k\text{th miRNA} \end{aligned} \quad (3)$$

and then each activity vector is normalized  $\mathbf{w}_k \leftarrow \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$ . As such, the activity of the miRNA will be deemed to be positive when its sequence-predicted targets are below their mean expression level. BayMiR considers a gene as a potential target of a miRNA if there is a complementary conserved match sites to the seed region of the miRNA.

### Processing mRNA expression data

The mRNA expression data were downloaded from the EMBL-EBI repository [80], available at [www.ebi.ac.uk/gxa/experiment/E-MTAB-62](http://www.ebi.ac.uk/gxa/experiment/E-MTAB-62). The data consists of 5,372 samples profiled on HG-U133A array platforms; As described in [80], the data were normalized and manually labeled into 369 biological groups covering a wide range of healthy/cancer tissues, conditions, and cell lines. We did the following processing on the retrieved expression data; all probe sets with no gene symbols were excluded. The samples belonging to each biological groups were averaged—the samples within one biological group are highly correlated ( $\rho > 0.85$ ). An upper/lower threshold defined by  $l_{th} = Q_2 - 1.5(Q_4 - Q_2)$  and  $u_{th} = Q_4 + 1.5(Q_4 - Q_2)$  respectively, when  $Q_2$  and  $Q_4$  represent the second and forth quartiles, were specified to detect and modify the extreme outliers. The outliers were then replaced with  $l_{th}$  or  $u_{th}$ . The gene symbol list in both expression and sequence datasets were updated based on the latest release of the HUGO Gene Nomenclature Committee (HGNC) (Feb.2012) to have consistent gene symbols.

### MiRNA-mRNA interaction analysis

We downloaded the list of 19,055 protein coding gene symbols from HGNC database and the list of 1,537 miRNA IDs from MiRbase V.19. We then built seven  $19,055 \times 1,532$  binary connectivity matrices based on the mRNA-miRNA interactions given by: Targetscan V6.1, [7] and TarBase [55]. All miRNAs are grouped into 1,251 miRNA families as defined by TargetScan—miRNAs sharing the same seed region. Conserved target sites are also retrieved from the TargetScan repository.

### Enrichment analysis

Gene ontology biological process (GO-BP) annotations were downloaded from the Gene Ontology Website on April 15th 2012. The file contains 14,000 annotations for 15,000 genes. The enrichment analysis was performed using Fisher Exact test. The test was performed on BayMiR predicted targets of each of miRNA families. The enrichment pvalues were corrected using Benjamini-Hochberg test [81] and a FDR cutoff equal to 0.1 was chosen to selected significant enrichment categories. The KEGG enrichment analysis carried out in a similar manner; The list of 253 KEGG human pathways were with associated genes downloaded from <http://www.genome.jp/kegg/>;

; Fisher exact test was used to find enriched pathways for BayMiR targets of all miRNA families.

### Availability of BayMiR and supporting data

The code for BayMiR is available at [morrislab.med.utoronto.ca/BayMiR](http://morrislab.med.utoronto.ca/BayMiR). package includes scripts and instructions to re-generate BayMiR scores from the "E-MTAB-62" file and sequence information, however, a pre-computed version of the BayMiR scores are also uploaded.

### Conclusions

We developed BayMiR, a new computational method for predicting the target mRNAs of miRNAs. BayMiR applies a large number of mRNA expression profiles and successfully identifies mRNA targets and miRNA activities without using miRNA expression data. We also showed that gene expression variability can be used to predict miRNA targets. Our analysis revealed the importance of miRNA target sites at 3' UTR near to the poly (A) tails. The BayMiR package is publicly available and can be applied to any mRNA expression datasets.

### Additional files

**Additional file 1: Figure S1.** Cumulative distribution of scores for the validated targets. Validated targets are assigned higher BayMiR scores and gene variation scores compared to the other putative targets. Shown are the cumulative distributions of BayMiR (left plot) and gene variation scores (right plot) scores for validated targets (blue) and all putative targets (red).

**Additional file 2: Figure S2.** Comparing BayMiR and Cometa. BayMiR high scoring targets are more down-regulated in miRNA over-expression assays than Cometa high scoring targets. The cumulative distribution of log-fold change for high-scoring mRNAs; blue, red, and black represent graphs associated with BayMiR, gene variation, and Cometa.

**Additional file 3: Table S1.** Excel file. Enriched GO-BP terms

**Additional file 4: Table S2.** Excel file. Enriched KEGG terms

**Additional file 5: Table S3.** Validated KEGG pathways. List of miRNAs with proposed functions found in our enriched KEGG list; the third column gives the Pubmed IDs of the references.

**Additional file 6: Figure S3.** KEGG "Pathways in cancer": 68 targets of 10 miRNAs are involved in the pathway (red boxes). 38 genes targeted by the other miRNAs are colored in yellow; and 62 genes involved in the pathway were excluded from the BayMiR target list since their expression variabilities across arrays were very low (white boxes). The miRNA family IDs: miR-17/17-5p/20ab/20b-5p/93/106ab/427/518a-3p/519d,miR-548ah/3609,miR-4729,miR-203,miR-548p,miR-3647-3p,miR-300/381/539-3p,miR-142-5p,miR-545,miR-125a-5p/125b-5p/351/670/4319'.

**Additional file 7: Table S4.** Excel file: miRNA expression data retrieved from the mimiRNA repository.

**Additional file 8: Figure S7.** Blue: the position contribution scores of miRNA-mRNA pairs whose BayMiR scores  $> median_{BayMiR scores}$ . Red: the position contribution scores of miRNA-mRNA pairs whose BayMiR scores  $< median_{BayMiR scores}$ .

**Additional file 9: Figure S4.** The 3' UTR of mRNAs harbor many conserved seed matches. Shown is the cumulative distribution of number of seed matches in the 3'UTR of 14,816 mRNA transcripts with at least one miRNA seed match.

**Additional file 10: Figure S5.** Example of combinatorial regulation masking inverse correlation. Shown in green is the expression level of a target gene and in red the expression levels of three targeting miRNAs. The negative correlation of each individual miRNAs with the target is insignificant, but when considered together they explain perfectly the down-regulation impact of miRNAs.

**Additional file 11: Figure S6.** Gene expression variability increases as the number of target sites increases in the 3' UTR of genes. (top) miRNA targets have high expression variation. (bottom) Red and blue demonstrate the cumulative distributions of genes whose variance is larger than median and 75<sup>th</sup> percentile, respectively. Dark: cumulative distribution of variances corresponding to all genes.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

Conceived and designed the experiments: MHR QM WW. Performed the experiments: MHR. Analyzed the data: MHR QM. Wrote the paper: MHR QM. All authors read and approved the final manuscript.

### Acknowledgements

This research was supported by Natural Science and Engineering Research Council grants to QM and WW. MHR was partially supported by an Ontario Graduate Scholarship.

### Author details

<sup>1</sup>Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada. <sup>2</sup>Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. <sup>3</sup>Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. <sup>4</sup>Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada.

Received: 8 April 2013 Accepted: 22 July 2013

Published: 30 August 2013

### References

- Bartel D: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136**(2):215–233.
- John B, Enright A, Aravin A, Tuschl T, Sander C, Marks D, et al.: **Human microRNA targets.** *PLoS Biol* 2004, **2**(11):e363.
- Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM, Zhang GZ: **Biological functions of microRNAs: a review.** *J Physiol Biochem* 2011, **67**:129–139.
- Ambros V: **The functions of animal microRNAs.** *Nature* 2004, **431**(7006):350–355.
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S, et al.: **The nuclear RNase III Drosha initiates microRNA processing.** *Nature* 2003, **425**(6956):415–419.
- Amere S, Martinez J, Schroeder R: **Molecular basis for target RNA recognition and cleavage by human RISC.** *Cell* 2007, **130**:101–112.
- Lewis B, Shih I, et al.: **Prediction of mammalian microRNA targets.** *Cell* 2003, **115**(7):787–798.
- Grimson A, Farh K, Johnston W, Garrett-Engele P, Lim L, Bartel D: **MicroRNA targeting specificity in mammals: determinants beyond seed pairing.** *Mol Cell* 2007, **27**:91–105.
- Betel D, Koppal A, Agius P, Sander C, Leslie C: **Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites.** *Genome Biol* 2010, **11**(8):R90.
- Khorshid M, Haussler J, Zavolan M, van Nimwegen E: **A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets.** *Nat Methods* 2013, **10**:253–255.
- Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R: **Fast and effective prediction of microRNA/target duplexes.** *Rna* 2004, **10**(10):1507.
- Friedman RC, Farh KK, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs.** *Genome Res* 2009, **19**:92–105.
- Nielsen C, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge C: **Determinants of targeting by endogenous and exogenous microRNAs and siRNAs.** *Rna* 2007, **13**(11):1894.
- Gaidatzis D, Van Nimwegen E, Haussler J, Zavolan M: **Inference of miRNA targets using evolutionary conservation and pathway analysis.** *BMC Bioinformatics* 2007, **8**:69.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: **The role of site accessibility in microRNA target recognition.** *Nature Genet* 2007, **39**(10):1278–1284.
- Tafer H, Ameres S, Obernosterer G, Gebeshuber C, Schroeder R, Martinez J, Hofacker I: **The impact of target site accessibility on the design of effective siRNAs.** *Nat Biotechnol* 2008, **26**(5):578–583.
- Majoros W, Ohler U: **Spatial preferences of microRNA targets in 3' untranslated regions.** *BMC Genomics* 2007, **8**:152.
- Garcia D, Baek D, Shin C, Bell G, Grimson A, Bartel D: **Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs.** *Nat Struct Mol Biol* 2011, **18**(10):1139–1146.
- Arvey A, Larsson E, Sander C, Leslie C, Marks D: **Target mRNA abundance dilutes microRNA and siRNA activity.** *Mol Syst Biol* 2010, **6**:220–225.
- Ritchie W, Flamant S, Rasko J: **Predicting microRNA targets and functions: traps for the unwary.** *Nat Methods* 2009, **6**(6):397–398.
- Barbato C, Arisi I, Frizzo M, Brandi R, Da Sacco L, Masotti A: **Computational challenges in miRNA target predictions: to be or not to be a true target?** *J Biomed Biotechnol* 2009, **1**:150–157.
- Saito T, Sætrom P: **MicroRNAs-targeting and target prediction.** *New Biotechnol* 2010, **27**(3):243–249.
- Hammell M: **Computational methods to identify miRNA targets.** In *Seminars in Cell & Developmental Biology*. Elsevier; 2010.
- Alexiou P, Maragakis M, Papadopoulos G, Reczko M, Hatzigeorgiou A: **Lost in translation: an assessment and perspective for computational microRNA target identification.** *Bioinformatics* 2009, **25**(23):3049–3055.
- Min H, Yoon S: **Got target?: computational methods for microRNA target prediction and their extension.** *Exp Mol Med* 2010, **42**(4):233.
- Guo H, Ingolia N, Weissman J, Bartel D: **Mammalian microRNAs predominantly act to decrease target mRNA levels.** *Nature* 2010, **466**(7308):835–840.
- Mukherji S, Ebert M, Zheng G, Tsang J, Sharp P, van Oudenaarden A: **MicroRNAs can generate thresholds in target gene expression.** *Nat Genet* 2011, **43**(9):854–859.
- Lim L, Lau N, Garrett-Engele P, Grimson A, Schelter J, Castle J, Bartel D, Linsley P, Johnson J: **Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.** *Nature* 2005, **433**(7027):769–773.
- Sood P, Krek A, Zavolan M, Macino G, Rajewsky N: **Cell-type-specific signatures of microRNAs on target mRNA expression.** *Proc Natl Acad Sci USA* 2006, **103**(8):2746.
- Filipowicz W, Bhattacharyya S, Sonenberg N: **Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?** *Nat Rev Genet* 2008, **9**(2):102–114.
- Baek D, Villén J, Shin C, Camargo F, Gygi S, Bartel D: **The impact of microRNAs on protein output.** *Nature* 2008, **455**(7209):64–71.
- Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N: **Widespread changes in protein synthesis induced by microRNAs.** *Nature* 2008, **455**(7209):58–63.
- Humphreys D, Westman B, Martin D, Preiss T: **MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly (A) tail function.** *Proc Natl Acad Sci USA* 2005, **102**(47):16961.
- Khan A, Betel D, Miller M, Sander C, Leslie C, Marks D: **Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs.** *Nat Biotechnol* 2009, **27**(6):549–555.
- Vivek J, David M, Yee Y: **Identification of microRNA-mRNA modules using microarray data.** *BMC Genomics* 2012, **13**.
- Liu B, Liu L, Tsykin A, Goodall G, Green J, Zhu M, Kim C, Li J: **Identifying functional miRNA-mRNA regulatory modules with correspondence latent dirichlet allocation.** *Bioinformatics* 2010, **26**(24):3105–3111.
- Sales G, Coppe A, Bisognin A, Biasiolo M, Bortoluzzi S, Romualdi C: **MAGIA, a web-based tool for miRNA and genes integrated analysis.** *Nucleic Acids Res* 2010, **38**(suppl 2):W352—W359.
- Yu-Ping W, Kuo-Bin L: **Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data.** *BMC Genomics* 2010, **11**.

39. Jayaswal V, Lutherborrow M, Ma D, Yang Y: **Identification of microRNAs with regulatory potential using a matched microRNA-mRNA time-course data.** *Nucleic Acids Res* 2009, **37**(8):e60–e60.

40. Ruike Y, Ichimura A, Tsuchiya S, Shimizu K, Kunimoto R, Okuno Y, Tsujimoto G: **Global correlation analysis for micro-RNA and mRNA expression profiles in human cell lines.** *J Human Genet* 2008, **53**(6):515–523.

41. Li X, Gill R, Cooper N, Yoo J, Datta S: **Modeling microRNA-mRNA interactions using PLS regression in human colon cancer.** *BMC Med Genom* 2011, **4**:44.

42. Muniategui A, Nogales-Cadenas R, Vázquez M, Aranguren X, Agirre X, Lutun A, Prosper F, Pascual-Montano A, Rubio A: **Quantification of miRNA-mRNA interactions.** *PLoS one* 2012, **7**(2):e30766.

43. Huang G, Athanassiou C, Benos P: **mirConnX: condition-specific mRNA-microRNA network integrator.** *Nucleic Acids Res* 2011, **39**(suppl 2):W416–W423.

44. Nam S, Li M, Choi K, Balch C, Kim S, Nephew K: **MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression.** *Nucleic Acids Res* 2009, **37**(suppl 2):W356–W362.

45. Wuchty S, Arjona D, Li A, Kotliarov Y, Walling J, Ahn S, Zhang A, Maric D, Anolik R, Zenklusen J, et al.: **Prediction of associations between microRNAs and gene expression in glioma biology.** *PLoS One* 2011, **6**(2):e14681.

46. Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, Hughes TR, Blencowe BJ, Frey BJ, Morris QD: **Using expression profiling data to identify human microRNA target.** *Nat Methods* 2007, **4**:1045–1049.

47. Huang J, Morris Q, Frey B: **Detecting microRNA targets by linking sequence, microRNA and gene expression data.** In *Research in Computational Molecular Biology*; Springer: 2006:114–129.

48. Huang J, Frey B, Morris Q: **Comparing sequence and expression data.** In *Pacific Symposium on Biocomputing, Volume 13*; 2008:52–63.

49. van Dongen S, Abreu-Goodger C, Enright A: **Detecting microRNA binding and siRNA off-target effects from expression data.** *Nat Methods* 2008, **5**(12):1023–1025.

50. Gennarino VA, D'Angelo G, Dharmalingam G, Fernandez S, Russolillo G, Sangres R, Mutarelli M, Belcastro V, Ballabio A, Verde P, et al.: **Identification of microRNA-regulated gene networks by expression analysis of target genes.** *Genome Res* 2012, **22**(6):1163–1172.

51. Gennarino VA, Sardiello M, Avellino R, Meola N, Maselli V, Anand S, Cutilli L, Ballabio A, Banfi S: **MicroRNA target prediction by expression analysis of host genes.** *Genome Res* 2008, **19**:481–490.

52. Peter M: **Targeting of mRNAs by multiple miRNAs: the next step.** *Oncogene* 2010, **29**(15):2161–2164.

53. Krek A, Grun D, Poy M, Wolf R, Rosenberg L, Epstein E, MacMenamin P, da Piedade I, Gunsalus K, Stoffel M, et al.: **Combinatorial microRNA target predictions.** *Nat Genet* 2005, **37**(5):495–500.

54. Friedman J, Hastie T, Tibshirani R: **Regularization paths for generalized linear models via coordinate descent.** *J Stat Soft* 2010, **33**:1.

55. Papadopoulos G, Reczko M, Simossis V, Sethupathy P, Hatzigeorgiou A: **The database of experimentally supported targets: a functional update of TarBase.** *Nucleic Acids Res* 2009, **37**(suppl 1):D155–D158.

56. Ulitsky I, Laurent L, Shamir R: **Towards computational prediction of microRNA function and activity.** *Nucleic Acids Res* 2010, **38**(15):e160–e160.

57. Huang JC, Morris QD, Frey BJ: **Bayesian inference of microRNA targets from sequence and expression data.** *J Comput Biol* 2007, **14**:550–563.

58. Volinia S, Calin G, Liu C, Ambros S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, et al.: **A microRNA expression signature of human solid tumors defines cancer gene targets.** *Proc Natl Acad Sci USA* 2006, **103**(7):2257–2261.

59. Uren A, Kool J, Matentzoglou K, De Ridder J, Mattison J, Van Uitert M, Lagcher W, Sie D, Tanger E, Cox T, et al.: **Large-scale mutagenesis in ip19ARF1 and ip53I deficient mice identifies cancer genes and their collaborative networks.** *Cell* 2008, **133**(4):727–741.

60. Ritchie W, Flamant S, Rasko J: **mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets.** *Bioinformatics* 2010, **26**(2):223–227.

61. Funakoshi Y, Doi Y, Hosoda N, Uchida N, Osawa M, Shimada I, Tsujimoto M, Suzuki T, Katada T: **Hoshino Si: Mechanism of mRNA deadenylation: evidence for a molecular interplay between translation termination factor eRF3 and mRNA deadenylases.** *Genes Dev* 2007, **21**(23):3135–3148.

62. Doench JG, Sharp PA: **Specificity of microRNA target selection in translational repression.** *Genes Dev* 2004, **18**(5):504–511.

63. Radfar M, Wong W, Morris Q: **Computational prediction of intronic microRNA targets using host gene expression reveals novel regulatory mechanisms.** *PLoS One* 2011, **6**(6):e19312.

64. Cheng C, Li L: **Inferring microRNA activities by combining gene expression with microRNA target prediction.** *PLoS One* 2008, **3**(4):e1989.

65. Cheng C, Fu X, Alves P, Gerstein M, et al.: **mRNA expression profiles show differential regulatory effects of microRNAs between estrogen receptor-positive and estrogen receptor-negative breast cancer.** *Genome Biol* 2009, **10**(9):R90.

66. Liang Z, Zhou H, He Z, Zheng H, Wu J: **mirAct: a web tool for evaluating microRNA activity based on gene expression data.** *Nucleic acids Res* 2011, **39**(suppl 2):W139–W144.

67. Alexiou P, Maragakis M, Papadopoulos G, Simmossis V, Zhang L, Hatzigeorgiou A: **The DIANA-mirExTra web server: from gene expression data to microRNA function.** *PLoS One* 2010, **5**(2):e9171.

68. Le Brigand K, Robbe-Sermesant K, Mari B, Barbuy P: **MiRonTop: mining microRNAs targets across large scale gene expression studies.** *Bioinformatics* 2010, **26**(24):3131–3132.

69. Volinia S, Visone R, Galasso M, Rossi E, Croce C: **Identification of microRNA activity by Targets' Reverse EXpression.** *Bioinformatics* 2010, **26**:91–97.

70. Arora A, Simpson D: **Individual mRNA expression profiles reveal the effects of specific microRNAs.** *Genome Biol* 2008, **9**(5):R82.

71. Yu Z, Jian Z, Shen S, Purisima E, Wang E: **Global analysis of microRNA target gene expression reveals that miRNA targets are lower expressed in mature mouse and Drosophila tissues than in the embryos.** *Nucleic Acids Res* 2007, **35**:152–164.

72. Liang Z, Zhou H, Zheng H, Wu J, Liang Z, Zhou H, Zheng H, Wu J, et al.: **Expression levels of microRNAs are not associated with their regulatory activities.** *Biol Direct* 2011, **6**:1–4.

73. Jayaswal V, Lutherborrow M, Yang Y: **Measures of association for identifying MicroRNA-mRNA pairs of biological interest.** *PLoS one* 2012, **7**:e29612.

74. Monteyns A, Spengler R, Wan J, Tededor L, Lennox K, Xing Y, Davidson B: **Structure and activity of putative intronic miRNA promoters.** *RNA* 2010, **16**(3):495.

75. Ozsolak F, Poling L, Wang Z, Liu H, Liu X, Roeder R, Zhang X, Song J, Fisher D: **Chromatin structure analyses identify miRNA promoters.** *Genes Dev* 2008, **22**(22):3172.

76. Martinez N, Ow M, Reece-Hoyes J, Barrasa M, Ambros V, Walhout A: **Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity.** *Genome Res* 2008, **18**(12):2005.

77. Lu J, Clark A: **Impact of microRNA regulation on variation in human gene expression.** *Genome Res* 2012, **22**(7):1243–1254.

78. Friedman J, Hastie T, Höfling H, Tibshirani R: **Pathwise coordinate optimization.** *Ann Appl Stat* 2007, **1**(2):302–332.

79. Piriyapongsa J, Mariño-Ramírez L, Jordan I: **Origin and evolution of human microRNAs from transposable elements.** *Genetics* 2007, **176**(2):1323–1337.

80. Lukk M, Kapushesky M, Nikkilä J, Parkinson H, Goncalves A, Huber W, Ukkonen E, Brazma A: **A global map of human gene expression.** *Nature Biotechnol* 2010, **28**(4):322–324.

81. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc Ser B (Methodological)* 1995:289–300.

doi:10.1186/1471-2164-14-592

Cite this article as: Radfar et al.: BayMiR: inferring evidence for endogenous miRNA-induced gene repression from mRNA expression profiles. *BMC Genomics* 2013 **14**:592.